

An evolution-based model for designing chorismate mutase enzymes

Year: 2020 | Citations: 258 | Authors: W. P. Russ, M. Figliuzzi, Christian Stocker, Pierre Barrat-Charlaix, M. Socolich

Abstract

Learning from evolution Protein sequences contain information specifying their three-dimensional structure and function, and statistical analysis of families of sequences has been used to predict these properties. Building from sequence data, Russ et al. used statistical models that take into account conservation at amino acid positions and correlations in the evolution of pairs of amino acids to predict new artificial sequences that will have the properties of the protein family. For the chorismate mutase family of metabolic enzymes, the authors demonstrate experimentally that the artificial sequences display natural-like catalytic function. Because the models access an enormous space of diverse sequences, such evolution-based statistical approaches may guide the search for functional proteins with altered chemical activities. Science, this issue p. 440 An evolution-based, data-driven engineering process can build artificial functional enzymes. The rational design of enzymes is an important goal for both fundamental and practical reasons. Here, we describe a process to learn the constraints for specifying proteins purely from evolutionary sequence data, design and build libraries of synthetic genes, and test them for activity in vivo using a quantitative complementation assay. For chorismate mutase, a key enzyme in the biosynthesis of aromatic amino acids, we demonstrate the design of natural-like catalytic function with substantial sequence diversity. Further optimization focuses the generative model toward function in a specific genomic context. The data show that sequence-based statistical models suffice to specify proteins and provide access to an enormous space of functional sequences. This result provides a foundation for a general process for evolution-based design of artificial proteins.